

WHAT IS CLAIMED IS:

1. A computer-implemented method of logically representing relationships between data elements defined according to a first physical representation of data, comprising:

providing a logical representation of the data, the logical representation abstractly describing a second physical representation of the data, wherein the second physical representation of the data is generated from the first physical representation of the data;

on the basis of the relationships between the data elements defined according to the first physical representation of the data, determining corresponding relationships between corresponding data structures defined according to the second physical representation of the data;

generating logical relationships abstractly describing the determined corresponding relationships, each logical relationship defining a path between data structures of the second physical representation; and

associating the generated logical relationships with the logical representation of the data.

2. The method of claim 1, wherein the logical representation comprises a plurality of logical field specifications, and wherein associating comprises including the generated logical relationships with respective logical field specifications.

3. The method of claim 1, wherein the first physical representation of the data is a document in text-based markup language.

4. The method of claim 3, wherein the text-based markup language is one of the eXtended Markup Language (XML) and the MicroArray Gene Expression Markup Language (MAGE-ML).

5. The method of claim 1, wherein the second physical representation is a relational representation.

6. The method of claim 5, wherein each data structure is a table of the relational representation.

7. The method of claim 1, wherein the first physical representation is a hierarchical representation and the second physical representation is a relational representation.

8. The method of claim 7, wherein the hierarchical representation is the eXtended Markup Language (XML).

9. The method of claim 7, wherein each data structure is a table of the relational representation.

10. The method of claim 1, further comprising removing any redundant determined corresponding relationships before generating the logical relationships.

11. A computer-implemented method of logically representing relationships between data elements defined according to a first physical representation of data, comprising:

- generating a second physical representation of the data from the first physical representation;

- generating a logical representation of the data as represented according to the second physical representation, the logical representation abstractly describing the second physical representation of the data;

- on the basis of the relationships between the data elements defined according to the first physical representation of the data, determining corresponding relationships between corresponding data structures defined according to the second physical representation of the data;

- generating logical relationships abstractly describing the determined corresponding relationships; and

including the generated logical relationships with the logical representation; wherein each of the generated logical relationships describes a path for traversing the second physical representation from a first data structure to a second data structure when processing a query requesting information related to the first and second data structures.

12. The method of claim 11, wherein the first physical representation of the data is a document in text-based markup language.

13. The method of claim 12, wherein the text-based markup language is one of the eXtended Markup Language (XML) and the MicroArray Gene Expression Markup Language (MAGE-ML).

14. The method of claim 11, wherein the second physical representation is a relational representation.

15. The method of claim 14, wherein each data structure is a table of the relational representation.

16. The method of claim 11, further comprising removing any redundant determined corresponding relationships before generating the logical relationships.

17. A computer-implemented method of logically representing relationships between data elements described in an eXtended Markup Language (XML) document, comprising:

retrieving a relational database schema for a plurality of data structures, each data structure corresponding to one of the data elements;

retrieving a logical representation abstractly describing the relational database schema;

determining the relationships between the data elements from the XML document;

on the basis of the determined relationships, determining corresponding relationships between corresponding data structures defined according to the relational database schema;

generating logical relationships abstractly describing the determined corresponding relationships; and

including the generated logical relationships with the logical representation; wherein each of the generated logical relationships describes a path for traversing a relational database constructed according to the relational database schema from a first data structure to a second data structure when processing a query requesting information related to the first and second data structures.

18. A computer-implemented method of querying physical data logically represented by a data abstraction model, wherein the physical data being queried is contained in data structures generated from a data source having a different schema from the data structures containing the physical data being queried, comprising:

receiving an abstract query comprising logical fields and corresponding values, wherein each of the logical fields is defined in the data abstraction model and wherein one or more of the logical fields are result fields to be returned by execution of the abstract query; and

transforming the abstract query into an executable query capable of being executed against the physical data; wherein the transforming is done using the data abstraction model and wherein the data abstraction model defines a specific path for traversing the data structures containing the physical data to reach the one or more result fields.

19. A computer-readable medium containing a program which, when executed by a processor, performs a process of logically representing relationships between data elements defined according to a first physical representation of data, the process comprising:

retrieving a logical representation of the data, the logical representation abstractly describing a second physical representation of the data, wherein the

second physical representation of the data is generated from the first physical representation of the data;

on the basis of the relationships between the data elements defined according to the first physical representation of the data, determining corresponding relationships between corresponding data structures defined according to the second physical representation of the data;

generating logical relationships abstractly describing the determined corresponding relationships, each logical relationship defining a path between data structures of the second physical representation; and

associating the generated logical relationships with the logical representation of the data.

20. The computer-readable medium of claim 19, wherein the logical representation comprises a plurality of logical field specifications, and wherein associating comprises including the generated logical relationships with respective logical field specifications.

21. The computer-readable medium of claim 19, wherein the first physical representation of the data is a document in text-based markup language.

22. The computer-readable medium of claim 21, wherein the text-based markup language is one of the eXtended Markup Language (XML) and the MicroArray Gene Expression Markup Language (MAGE-ML).

23. The computer-readable medium of claim 19, wherein the second physical representation is a relational representation.

24. The computer-readable medium of claim 23, wherein each data structure is a table of the relational representation.

25. The computer-readable medium of claim 19, wherein the first physical representation is a hierarchical representation and the second physical representation is a relational representation.
26. The computer-readable medium of claim 25, wherein the hierarchical representation is the eXtended Markup Language (XML).
27. The computer-readable medium of claim 25, wherein each data structure is a table of the relational representation.
28. The computer-readable medium of claim 19, wherein the process further comprises:
removing any redundant determined corresponding relationships before generating the logical relationships.
29. A computer-readable medium containing a program which, when executed by a processor, performs a process of logically representing relationships between data elements defined according to a first physical representation of data, the process comprising:
generating a second physical representation of the data from the first physical representation;
generating a logical representation of the data as represented according to the second physical representation, the logical representation abstractly describing the second physical representation of the data;
on the basis of the relationships between the data elements defined according to the first physical representation of the data, determining corresponding relationships between corresponding data structures defined according to the second physical representation of the data;
generating logical relationships abstractly describing the determined corresponding relationships; and
including the generated logical relationships with the logical representation; wherein each of the generated logical relationships describes a path for traversing

the second physical representation from a first data structure to a second data structure when processing a query requesting information related to the first and second data structures.

30. The computer-readable medium of claim 29, wherein the first physical representation of the data is a document in text-based markup language.

31. The computer-readable medium of claim 30, wherein the text-based markup language is one of the eXtended Markup Language (XML) and the MicroArray Gene Expression Markup Language (MAGE-ML).

32. The computer-readable medium of claim 29, wherein the second physical representation is a relational representation.

33. The computer-readable medium of claim 32, wherein each data structure is a table of the relational representation.

34. The computer-readable medium of claim 29, wherein the process further comprises:

removing any redundant determined corresponding relationships before generating the logical relationships.

35. A computer-readable medium containing a program which, when executed by a processor, performs a process of logically representing relationships between data elements described in an eXtended Markup Language (XML) document, the process comprising:

retrieving a relational database schema for a plurality of data structures, each data structure corresponding to one of the data elements;

retrieving a logical representation abstractly describing the relational database schema;

determining the relationships between the data elements from the XML document;

on the basis of the determined relationships, determining corresponding relationships between corresponding data structures defined according to the relational database schema;

generating logical relationships abstractly describing the determined corresponding relationships; and

including the generated logical relationships with the logical representation; wherein each of the generated logical relationships describes a path for traversing a relational database constructed according to the relational database schema from a first data structure to a second data structure when processing a query requesting information related to the first and second data structures.

36. A computer-readable medium containing a program which, when executed by a processor, performs a process of querying physical data logically represented by a data abstraction model, wherein the physical data being queried is contained in data structures generated from a data source having a different schema from the data structures containing the physical data being queried, the process comprising:

receiving an abstract query comprising logical fields and corresponding values, wherein each of the logical fields is defined in the data abstraction model and wherein one or more of the logical fields are result fields to be returned by execution of the abstract query; and

transforming the abstract query into an executable query capable of being executed against the physical data; wherein the transforming is done using the data abstraction model and wherein the data abstraction model defines a specific path for traversing the data structures containing the physical data to reach the one or more result fields.

37. The computer-readable medium of claim 36, wherein the specific path is derived from relationships in the data source.

38. The computer-readable medium of claim 36, wherein the data source is a document in text-based markup language.

39. The computer-readable medium of claim 38, wherein the text-based markup language is one of the eXtended Markup Language (XML) and the MicroArray Gene Expression Markup Language (MAGE-ML).

40. The computer-readable medium of claim 36, wherein the data structures containing the physical data being queried are arranged according to a relational schema.

41. The computer-readable medium of claim 40, wherein each data structure containing physical data being queried is a database table according to the relational schema.

42. The computer-readable medium of claim 36, wherein the data source is arranged according to a hierarchical representation and the data structures containing the physical data being queried define a relational representation.

43. The computer-readable medium of claim 42, wherein the hierarchical representation is the eXtended Markup Language (XML).

44. A data structure residing in memory, comprising:
a plurality of logical field specifications, each abstractly describing at least one of a plurality of data structures defined according to a physical representation of data, wherein at least one of the plurality of logical field specifications includes one or more logical relationships algorithmically generated from relationship information describing relationships between the data represented according to another physical representation of the data, each logical relationship describing a path for traversing the physical representation of the data from a first data structure to a second data structure when processing a query requesting information related to the first and second data structures.